

Improving Facial Expression Recognition Using HOG with SVM and Modified Datasets Classified by Alexnet



Salar Jamal Abdulhameed Al-Atroshi^{*}, Abbas M. Ali[†]

Software and Informatics Engineering Department, College of Engineering, Salahaddin University-Erbil, Erbil 44002, Iraq

Corresponding Author Email: salar.atroshi@su.edu.krd

<https://doi.org/10.18280/ts.400429>

ABSTRACT

Received: 17 January 2023

Revised: 20 June 2023

Accepted: 4 July 2023

Available online: 31 August 2023

Keywords:

CNN, datasets, facial expression, hog, SVM

Facial expressions are one of the communication ways between humans and their behavior can be determined through their facial expressions. Recently computer technology has been used to identify the facial expressions of people in order to predict their intentions. It remains a challenge for facial expression recognition to extract discriminative features from training sets with few labels because most deep learning-based algorithms primarily rely on spatial information and large labels. In this paper, we propose a system to classify seven types of facial expressions (Angry, Sadness, Surprise, Happiness, Fear, Neutral, and Disgust) instead of six, as in most previous research. In the proposed system, machine learning algorithms with deep learning are used to increase classification accuracy based on removing some unimportant facial regions. The support vector machine (SVM) algorithm is trained to detect the eyes and mouth regions from the face depending on histogram-oriented gradient (HOG) which is used as a features extractor. Then, merge the eyes and mouth regions for each image to create a new form of an image. After that, five different types of images are generated from the merged image named (RGB, HSV, Gray, Binary, and YCbCr). The images are fed one by one into the convolution neural network (CNN) algorithm. Finally, the voting process is used to select the most predictive class. The proposed system has been tested on three different types of datasets (KDEF, JAFFE, and FER2013) and the prediction accuracy in the system has reached more than 98% in all used datasets. The conclusion is that eliminating unimportant regions impacts the results of the classification accuracy.

1. INTRODUCTION

Facial Expression Recognition (FER) is among the most efficient methods for including nonverbal information in human-behavior analysis because it provides some clues regarding emotional states and intentions [1]. For over three decades. Humans use facial expressions to interact with their emotional states and intentions. It represents one of the most potent, natural, and universal signals available [2, 3]. Due to its significant role in medical treatment [4], driver fatigue surveillance [5-7], sociable robotics [8-10], and many other human-computer interaction systems [11-13], automatic facial expression analysis has been the subject of several studies. Various facial expression recognition (FER) systems have been investigated in the field of machine learning and computer vision to encode expression information from facial representations. Ekman and Friesen [14] presented six fundamental expressions depending on cross-cultural research [15] as early as the twentieth century, denoting that humans similarly recognize certain basic expressions regardless of their culture. Fear, happiness, anger, disgust, surprise, and sadness are the most common facial expressions. Contempt was later added to the list of basic emotions [16].

Psychological characteristics which include facial expressions, hand gestures, voice, body parts movement, and variations in heartbeat and blood pressure can be used to categorize human emotions. The human face expresses itself through skeptical entities such as the mind, facial muscles,

condition, and surroundings. The method of presenting the human emotions or mental state by analyzing the motion of facial components, such as lip movement counts, eye blinks, thin and extended eyelids, skewed eyebrows, pulling down eyebrows, and creasing nose can be broadly defined as facial expression recognition [17]. The emotions expressed on the human face influence decisions and debates on various topics.

Artificial intelligence (AI) plays a critical role in developing FER software. Through machine learning algorithms, AI systems can be taught to identify patterns and features that correspond to specific emotions. This involves training the systems on large datasets of images and using this as a basis for building a model that can accurately identify FEs in new images. AI has become increasingly important in FER this is because AI algorithms process the ability to learn from data, allowing them to detect and analyze patterns in FEs quickly and accurately. It can also recognize which facial muscles are activated during certain emotions and can use this information to identify and interpret facial expressions [18]. FER has a number of important applications for different industries, from improving customer service experience systems. For example, AI-powered FER can help retailers to gauge how customers respond to different products, helping them to optimize sales strategies. Overall, the use of AI in FER has opened up new possibilities for technology, making it an invaluable tool for a wide range of industries. Its ability to accurately recognize and interpret FEs offers a host of benefits [19]. AI calculations, particularly deep neural organization, are capable of learning

complex provisions and grouping the extracted designs.

The main contributions of this paper are listed below:

1. We proposed a robust system for facial expression recognition which would be competitive with the performance of the other approaches.

2. The features of the available datasets are enhanced to increase the prediction accuracy.

3. To show the strength of the proposed system, we compared the proposed system with the traditional CNN model (Alexnet) used for facial expression recognition.

The rest sections of the paper are structured as follows: A related work is explained in section 2. Histogram of oriented gradient (HOG), support vector machine (SVM), and convolutional neural network (CNN) are briefly presented in sections 3, 4, and 5 respectively. Section 6 provides details about the proposed system and results and discussion are provided in section 7. Finally, conclusions are given in section 8.

2. RELATED WORK

Umer et al. [20] proposed a method to identify various kinds of expressions inside the human face region. The suggested system's execution is split into four modules. The results of experiments have been evidenced using three different datasets: CK+(seven expression classes), GENKI-4k (two expression classes), and KDEF (seven expression classes). Although this work was proposed for solving the overfitting problem, the results were not quite enough to be reliable for more applications.

Podder et al. [21] used real-time facial expressions to assist in representing patients' behavioral psychology. Unlike standard lightweight convolutional neural networks (CNNs), the proposed one employs a bypass connection to improve gradient propagation through the network, resulting in higher accuracy. CK+ and FER-2013 datasets were used to obtain 96.12% and 68.93% accuracies for the experiment, respectively. It is obvious that the classification accuracy of the FER-2013 dataset is low.

A Fuzzy optimized (CNN-RNN) method for facial expression recognition was proposed by Zhang and Tian [22]. The Fuzzy logic is used to reduce the suspicion in the activation map by recognizing the highly nonlinear relationship between the features. The results of experiments depending on the open datasets JAFFE, FER2013, and CK+ illustrate that the proposed Fuzzy optimized CNN-RNN method outperforms current popular techniques in the recognition of different facial expression datasets which gives an accuracy of 96.64%, 72.81%, and 99.22% respectively, but still the classification accuracy of the FER-2013 dataset is not quite enough.

Sadik et al. [23] suggested a model that has been implemented using the MobileNet model's Convolutional Neural Network (CNN) for autism recognition. The experimental results demonstrate that the MobileNet model using a transfer learning technique might achieve good results in the recognition task, with the highest validation accuracy of 89% and test accuracy of 87%.

González-Lozoya et al. [1] proposed an approach for recognizing facial expressions by employing features extracted with CNNs, making use of a pre-trained model in similar processes. The experimental results demonstrate that the FER technique can recognize the six basic expressions

with an accuracy of more than 92% when using five broadly utilized datasets. The classification accuracy is still low for six expressions.

Haque and Valles [24] proposed a model to teach young autistic children to recognize human facial expressions using image processing and computer vision. The research focuses primarily on the preliminary work on facial expression recognition using a deep convolutional neural network. The FER2013 database was used to train and test a deep CNN model. The results of the work still do not reach to be dependable.

3. HISTOGRAM OF ORIENTED GRADIENT (HOG)

The HOG feature is extracted and enumerated via the histogram of the gradient direction of the image's local area, and finally creates features that can efficiently extract facial expression features [25]. To improve performance, the HOG feature is counted on a dense image grid with uniform intervals, and overlapped local contrast normalization has been used [26]. The HOG is a feature descriptor that is utilized in image processing and computer vision for object detection [27].

The HOG features can be calculated by the following steps:

3.1 Gradient calculation

Gradient refers to the change in intensity or color of adjacent pixels in an image. In facial expression analysis, the gradient can be used to detect edges and contours of features like the eyes, mouth, and eyebrows. This information can be used to determine the direction and magnitude of movement in the facial expression.

First, the vertical and horizontal gradient maps are computed by convolving two Sobel filters and expression images. $[-1,0,1]^T$ is the vertical edge operator, and $[-1,0,1]^T$ is the horizontal edge operator. Gamma and smooth normalization operations, in particular, can be ignored [28].

3.2 Calculation of amplitude and direction

The amplitude and direction maps are computed in step 1 using the vertical and horizontal gradient maps. Assuming that dx and dy are the gradient values in the horizontal and vertical maps.

The amplitude and gradient of the pixel can be calculated using Eq. (1) and Eq. (2) [26].

$$\text{Magnitude} = \sqrt{(\text{dx})^2 + (\text{dy})^2} \quad (1)$$

$$\text{Orientation} = \tan^{-1} \left(\frac{\text{dy}}{\text{dx}} \right) \quad (2)$$

Magnitude refers to the strength or intensity of the gradient or the amount of change in color or intensity. In facial expression analysis, the magnitude can be used to measure the intensity of the expression. For example, a smile with a high magnitude would indicate a more positive and intense emotion compared to a smile with a lower magnitude.

Orientation, on the other hand, is the direction of the facial feature's movement. For example, the orientation of the eyebrows can indicate whether a person is surprised, angry, or sad. The orientation can be determined by tracking the movement of the eyebrow points over time.

In combination, magnitude and orientation information can be used to detect and classify facial expressions, which can have a range of applications. For example, this kind of analysis can be used in psychology to study emotions, in human behavior research to understand the impact of different stimuli on facial expressions, and in various human-computer interaction applications, such as gaming or virtual assistants.

Together, gradient and magnitude analysis can be used to detect and quantify facial expressions and emotions. They are important tools in the field of facial expression recognition and can be used in applications such as emotion detection in virtual assistants, video analytics, and autism therapy.

3.3 Unit quantization

Unit quantization for facial expression refers to the process of breaking down facial expressions into discrete units or categories. This allows for easier recognition and analysis of facial expressions, as each expression can be categorized based on specific units or codes. For example, the Facial Action Coding System (FACS) is a unit quantization system that breaks down facial expressions into specific muscle movements or "action units". By quantifying facial expressions, this information can be used in a variety of fields, including psychology, neuroscience, and even digital animation and robotics.

The gradient direction value range is 0 to 180, which is evenly divided into 9 intervals of 20 degrees each. The gradient amplitude is utilized as the projection weight [28].

3.4 Block normalization

Block normalization can be applied to HOG features to improve their robustness to lighting and pose variations in facial expression recognition. In HOG, block normalization involves dividing the image into small blocks and normalizing the gradient magnitude values within each block.

In most situations, unequal lighting could affect the gradient's amplitude, resulting in various value ranges, and local contrast normalization can improve the robustness because of lighting conditions and improve performance. Eq. (3) can be used to calculate the normalization process.

$$v = \sqrt{\frac{v}{(|v| + E)}} \quad (3)$$

where, v is the eigenvector before normalization, and E is the constant that causes the denominator to be non-zero [28].

4. SUPPORT VECTOR MACHINE (SVM) CLASSIFIER

Support Vector Machines (SVM) can be used for facial expression recognition. SVM is a supervised learning algorithm that is commonly used in pattern classification and regression analysis. In facial expression recognition, SVM is used to classify facial images into different categories based on the emotions they convey.

To use SVM for facial expression recognition, the first step is to prepare a dataset of images labeled with their corresponding facial expressions. The images can be preprocessed to extract features such as facial landmarks or texture information, which are then used in SVM training. In

this work, SVM has been used for detecting the mouth and eyes regions inside the face of the human.

Once the dataset is ready, the SVM algorithm is trained with the extracted features from each image as input and the corresponding facial expression as the output. The trained model can then be used to classify new facial images into different emotional categories for the eyes and mouth.

One of the advantages of SVM is that it is effective in handling high-dimensional datasets, such as those that arise in image analysis. Additionally, SVM is robust to noisy data and can produce good results even with limited training data.

SVM is a method that utilizes the hypothesis space of a linear function in a high-dimensional feature space and is trained by a learning algorithm based on the theory of optimization that applies a learning bias coming from the theory of statistical learning [29]. All over the world, the support vector machine is an effective component of machine learning research, also it is a binary classification algorithm that treats each data point as a k -dimensional vector. The SVM generates separation hyperplanes between samples from two classes (labeled -1 and +1) [27]. Figure 1 illustrates the process of SVM classification.

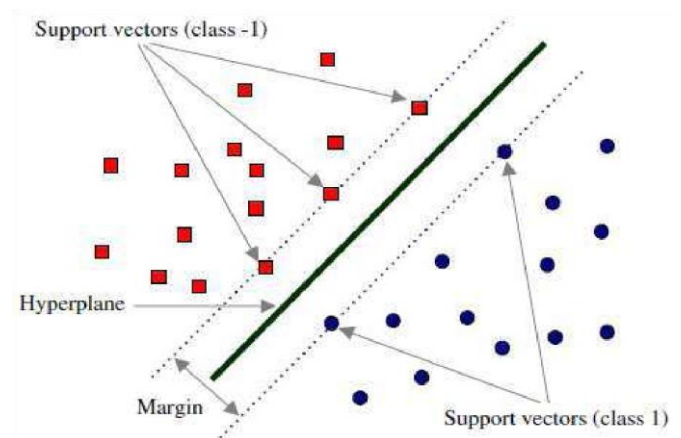


Figure 1. SVM classification [30]

In summary, SVM can be a useful tool for facial expression recognition since it can effectively classify facial images into different emotional categories based on the features extracted from the images like eyes and mouth.

5. CONVOLUTIONAL NEURAL NETWORK (CNN)

Convolutional neural networks are extremely much like artificial neural networks, except that the inputs are assumed to be images. Due to this assumption, researchers can encode specific properties into the CNN architecture. A standard Convolutional neural network architecture is made up of several layers that allow it to learn the hierarchical feature representation of the images. The existence or not of edges at specific orientations and locations in the image is generally represented by features in the first layer of representation. The second layer locates motifs by spotting edge arrangements, irrespective of slight differences in edge locations. The third layer would combine motifs into larger combinations that correlate to components of familiar objects, and the following layers could detect objects as an outcome of these combinations [31].

The convolutional neural network has three main layers:

5.1 Convolution layer

A CNN's core building unit is the convolution layer. The parameters of the convolution layer are made up of a collection of learnable filters. Every filter can be small in space but extends throughout the entire depth of the input image. By sliding the filter over the input image, a 2D map is created that contains the filter's responses at each spatial location. Each convolution layer includes a set of filters, and every filter generates a separate map. The output image is a stack of these depth-dimensional maps. Each entry in the output image can be construed as the neuron output that examines a small region of the input and shares parameters with neurons within the same feature map [31].

5.2 Pooling layer

The pooling layer is doing spatial dimension reduction. Adding a pooling layer between consecutive convolution layers regularly is a common practice. The pooling layer's function is to decrease the spatial size of the representation in order to reduce the number of parameters and computations in the network while also preventing overfitting. The pooling layer works autonomously on each depth slice of the input and spatially resizes it. The MAX pooling function is the most widely used pooling function, and it utilizes the maximum

value from each group of neurons in the previous layer to create a new neuron in the subsequent layer [31, 32].

5.3 Fully connection layer

As demonstrated in traditional neural networks, the fully connected layer is a layer of neurons with full connections to all the outputs of the previous layer. The convolution and pooling layers extract features from the input image, while the fully connected layer functions as a classifier, where a Soft-Max classifier or sigmoid is used to predict the input class label [31].

6. PROPOSED SYSTEM

One of the significant challenges facing this research is the difficulty of distinguishing between the expressions of human faces. In this paper, a proposed system that combines image processing, machine learning algorithm, and deep learning algorithm is utilized to increase prediction accuracy. Where used different types of image colors, histogram of orientated gradient (HOG) for features extraction, support vector machine (SVM) for eyes and mouth detection, and convolutional neural network for expression classification. Figure 2 illustrates the proposed system block diagram.

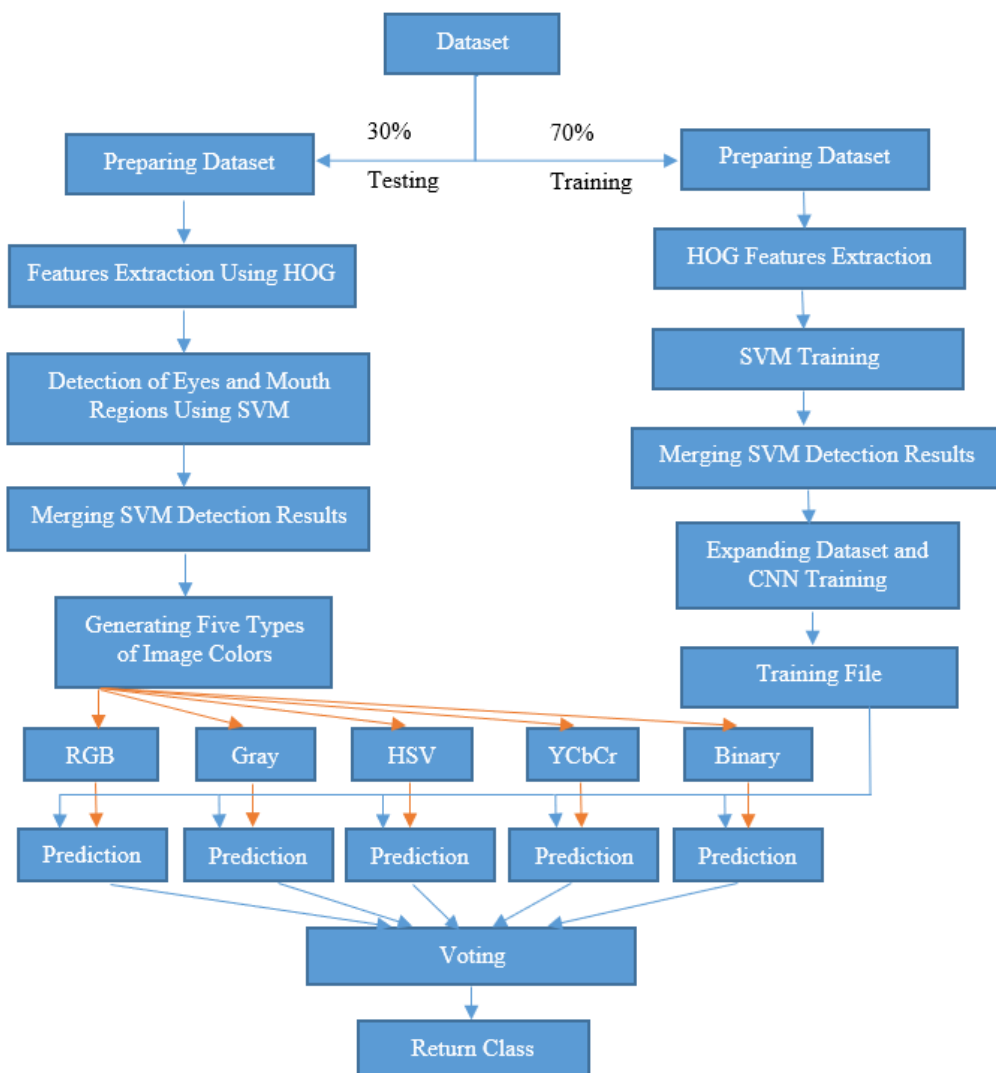


Figure 2. Proposed system block diagram

6.1 Training phase

The training phase consists of two steps of training which are:

a. SVM Training: The first step in the proposed system is preparing the dataset by generating three classes (eyes, mouths, and random class), then using HOG to extract features of each class to train SVM for detections.

The training process involves the following steps:

1. Data collection: Collect images of faces with different expressions and label them according to their corresponding categories (eyes, mouths, and random class).

2. Feature extraction: Extract a set of features from the image regions using HOG, such as the position and intensity of the eyes, and mouth.

3. Feature normalization: Normalize the features to have a mean of 0 and a standard deviation of 1 to improve the performance of the SVM model.

4. Model training: Train an SVM model using the labeled dataset with the extracted normalized features.

5. Hyperparameter tuning: Tune the hyperparameters of the SVM model, such as the regularization parameter, kernel function, and margin to improve the classification performance for the eyes and mouth regions.

6. Model evaluation: Evaluate the performance of the

trained SVM model on a test dataset by measuring its accuracy.

7. Deployment: Deploy the trained SVM model to classify new facial expression regions into their corresponding categories (eyes or mouth).

b. CNN Training: The second step is merging the results of SVM detection images, then expanding these results by generating five types of image colors to train the Alexnet model. Figure 3 demonstrates the Alexnet structure.

Training Alexnet involves the following steps:

1. Data preprocessing: Collect facial expression images and preprocess them by resizing and merging the results of SVM detection images, expanding these images to five different types of image colors, and normalizing them to a fixed size of $227 \times 227 \times 3$.

2. Feature extraction: Extract deep features from the preprocessed facial expression images using a pre-trained Alexnet model, which is designed to capture complex and abstract features from images.

3. Data splitting: Split the data into training and testing sets, with a ratio of 70:30.

4. Model training: Train the Alexnet for each facial expression category using the selected features from the training dataset.

5. Model evaluation: Evaluate the trained Alexnet on the testing dataset to measure its accuracy.

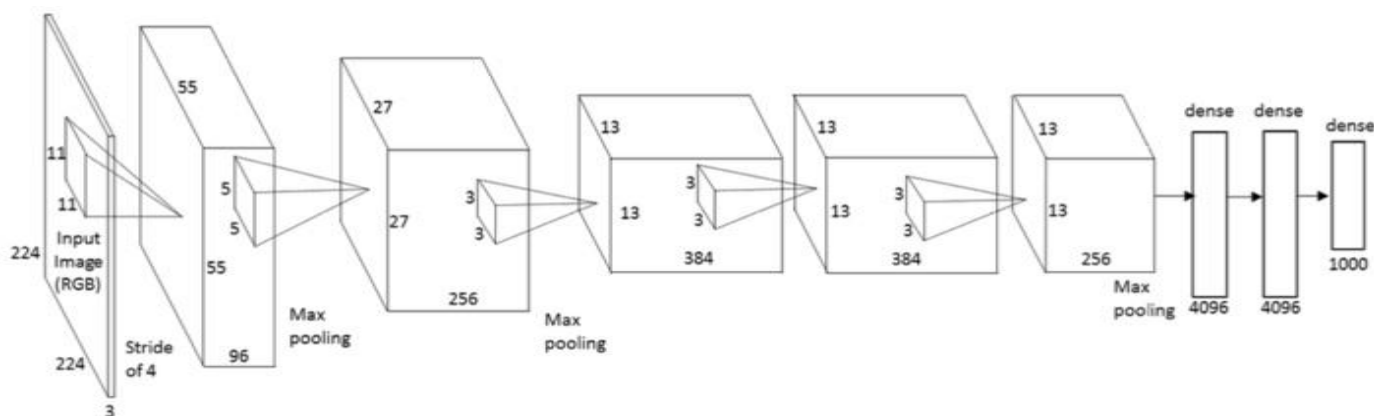


Figure 3. Alexnet structure [33]

6.2 Testing phase

The testing phase of Alexnet for facial expression recognition would involve the preparation of the test dataset. This dataset should contain facial images (eyes and mouth) of individuals displaying different emotions (happy, sad, angry, etc.), then loading a pre-trained model where Alexnet is a pre-trained model on ImageNet, so the next step is to load and set up the model for facial expression recognition.

The testing process is then tested on the test dataset to evaluate its performance. During the testing process, the input images are fed into the model and the predicted emotions are compared to the actual emotions. After that, the performance of the model is evaluated based on accuracy. In this work, the testing phase can be summarized by the following steps.

a. Eyes and Mouth Detection: at this step, we use a multiscale window and HOG to extract the features of each window and then categorize them into one of three classes (Eyes, Mouths, and Random) by using SVM.

b. Cropping Step: at this step, the proposed system crops

and saves only (Eyes, and Mouth) detections and neglects random class detection.

c. Merging Step: at this step, cropped eyes image and cropped mouth image are merged to form a new image.

d. Image Colors Generation: at this step, the proposed approach generates five types of image colors (RGB, Gray, HSV, YCbCr, and Binary).

e. Prediction: at this step, the proposed system predicts each of the five images and returns the target class then enters the results in an array.

f. Voting: Finally, the proposed system chooses the most repeated class in the prediction array.

7. RESULTS AND DISCUSSION

Many practical experiments are conducted on the proposed system where the performance of the proposed system is measured by using three datasets which are (KDEF, JAFFE, and FER2013). Figure 4 shows samples of these datasets.

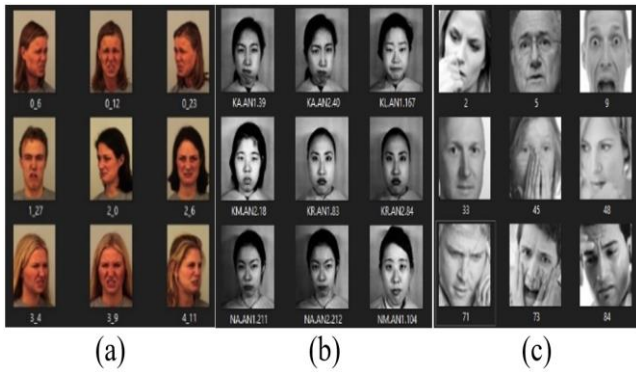
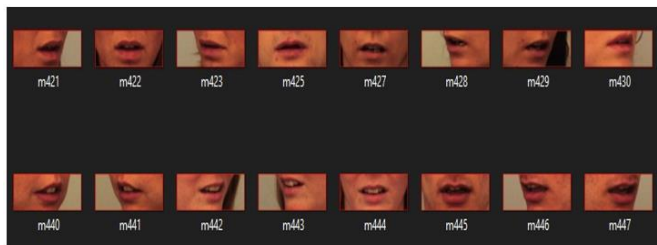


Figure 4. Facial expression datasets (a) KDEF (b) JAFFE (c) FER2013



(a)



(b)

Figure 5. (a) Results of SVM eyes detection (b) Results of SVM mouth detection

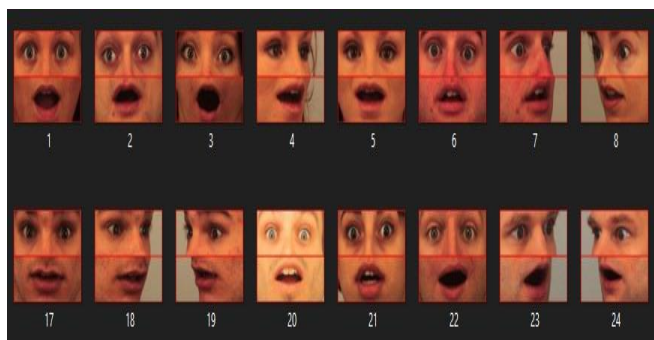


Figure 6. Merging process results

Table 1. Summary of SVM detection results

| Dataset | Total Number of Images | Number of True Detection | Number of False Detection | Accuracy |
|---------|------------------------|--------------------------|---------------------------|----------|
| KDEF | 2,940 | 2,931 | 9 | 99.69% |
| JAFFE | 210 | 209 | 1 | 99.52% |
| FER2013 | 22,197 | 22,094 | 103 | 99.54% |

The using of HOG with the SVM classifier has shown highly accurate results in eyes and mouth detections, where the

results of accuracy reached more than 99%. Figure 5 illustrates the results of eye and mouth detections. Table 1 illustrates the accuracy of detections.

The next step is merging the SVM results to generate a new image. Figure 6 illustrates the merging process results.

After that, the proposed system generates (RGB, HSV, Gray, YCbCr, and Binary) images. Figure 7 illustrates the results of generated images.

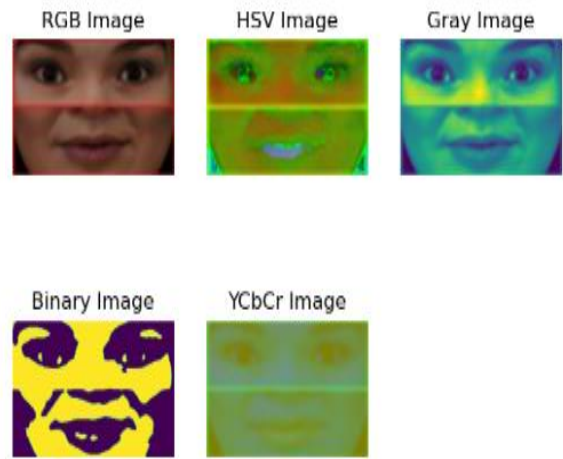


Figure 7. Results of images colors generation

Different datasets are trained by the same CNN model (Alexnet), and each dataset contains a different number of images. The accuracy obtained from training the traditional CNN with original datasets is low for facial expression classification. Table 2 illustrates the results of CNN before using the proposed system.

Whereas training these datasets after using the proposed system, the accuracy of training and validation is increased significantly. Table 3 illustrates the results after implementing the proposed system.

Depending on the results shown in Table 2 and Table 3, we note that the accuracy of the prediction raised dramatically after the implementation of the proposed system. Finally, employing the voting mechanism has a big role to improve the prediction. Table 4 illustrates a summary of all results.

The accuracy of the proposed system is important. It reached more than 99% for SVM used to detect the eyes and mouth regions and for the whole proposed system is more than 98% which is used for the classification of the emotions in all used datasets. These results determine how well the proposed model is reliably performing facial expression recognition. It performs well on facial recognition tasks based on our proposed features of the eyes and mouth regions. The performance of the proposed model outperforms the other works listed in the literature.

Expanding the dataset is overcome the overfitting problem since it enlarged the dataset, so it improved the overall performance of the model.

Facial Expressions can vary based on scene changes and lighting conditions. Alexnet is susceptible to these changes, but adding more color spaces to the dataset can improve its ability to detect facial expressions.

The facial expressions learned by Alexnet can transfer to other recognition tasks like emotion recognition, facial detection, or facial landmark recognition. In conclusion, Alexnet is a strong baseline model for facial expression recognition. It is robust and generalizes well.

Table 2. CNN results before using the proposed system

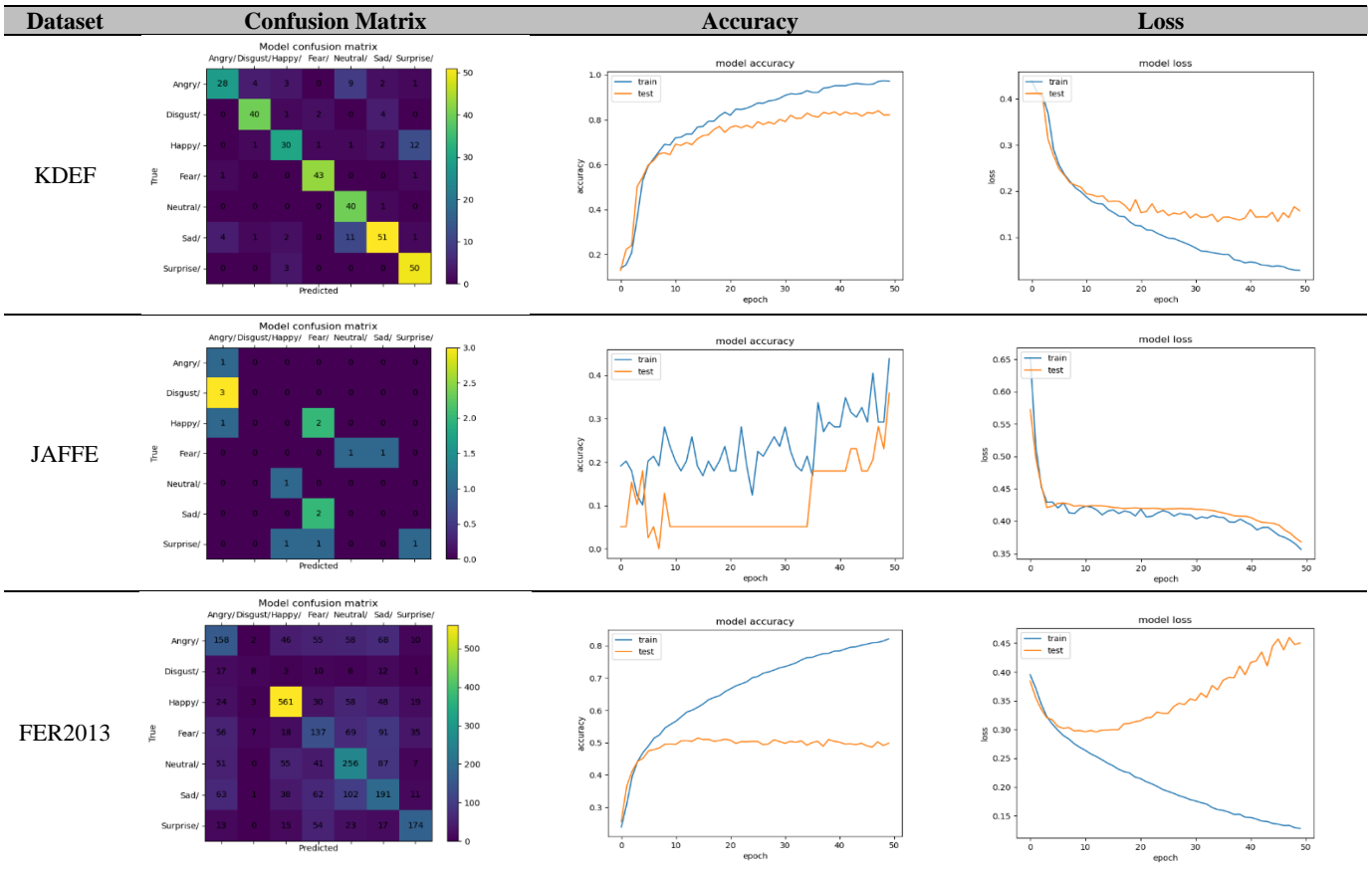


Table 3. CNN results after using the proposed system

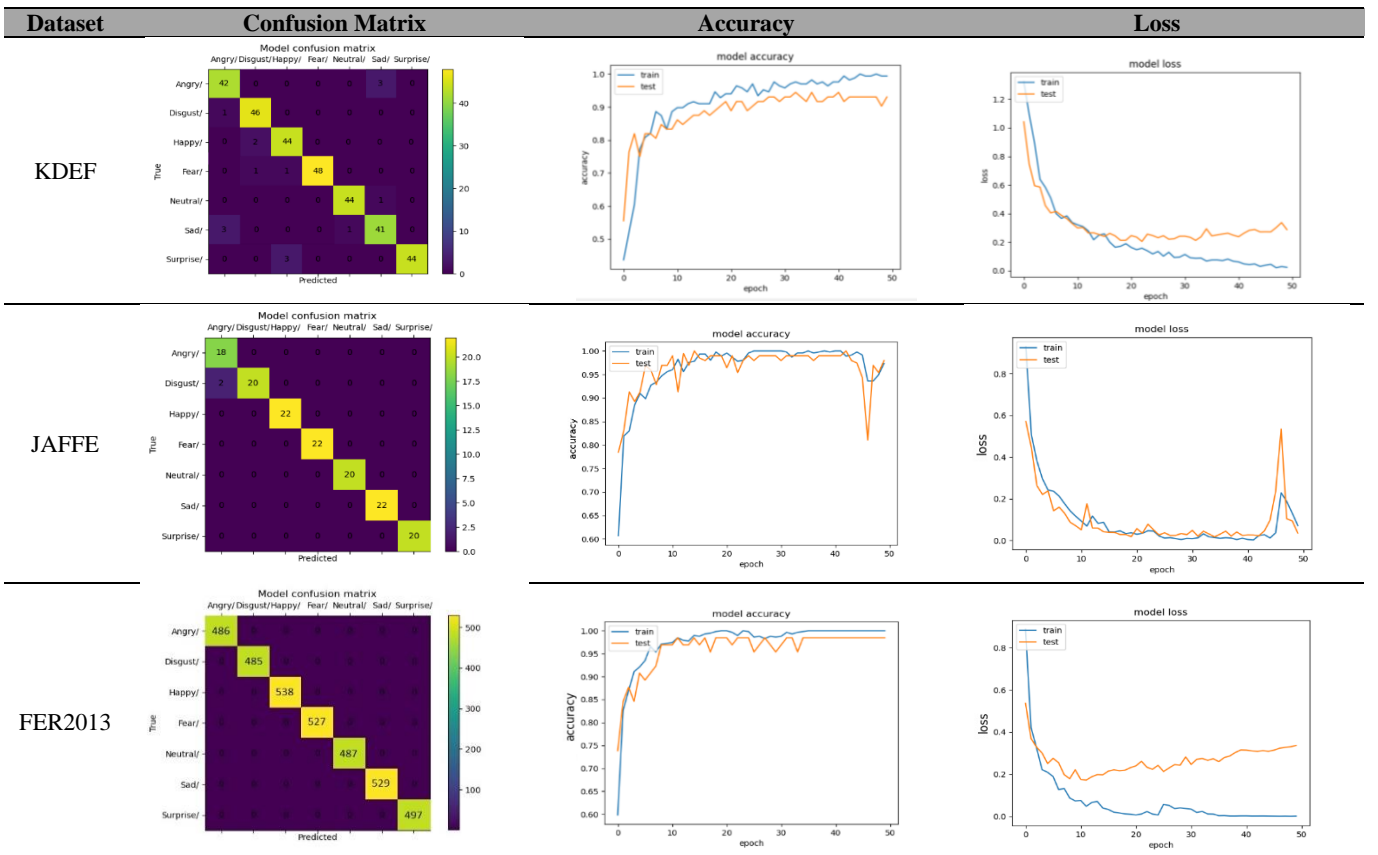


Table 4. Summary of all results

| Algorithm | Dataset | Training Accuracy | Training Loss | Testing Accuracy | Testing Loss |
|---------------------------------------------------------------|---------|-------------------|---------------|------------------|--------------|
| CNN | KDEF | 0.9953 | 0.0254 | 0.8241 | 0.1987 |
| | JAFFE | 0.4576 | 0.3547 | 0.3987 | 0.3694 |
| | FER2013 | 0.8125 | 0.1457 | 0.5021 | 0.4574 |
| CNN, Machine Learning, and Image Processing | KDEF | 0.9951 | 0.1471 | 0.9425 | 0.3541 |
| | JAFFE | 0.9741 | 0.1089 | 0.9789 | 0.0931 |
| | FER2013 | 1.000 | 0.0014 | 0.9778 | 0.3740 |
| CNN, Machine Learning, and Image Processing with Voting | KDEF | - | - | 0.9847 | 0.0471 |
| | JAFFE | - | - | 0.9899 | 0.0321 |
| | FER2013 | - | - | 0.9876 | 0.0337 |

8. CONCLUSIONS

The aim of this study designs an efficient model to recognize human facial expressions. The proposed system is developed by using image processing to expand the databases, machine learning to locate the required facial expressions regions which are eyes and mouth, and CNN for classification. In this research, three databases are used, each including seven facial expressions. From the results of the proposed system, we find that the performance of the model after expanding the dataset is better than the performance of the original dataset before expanding. on the other hand, we note that deducting the eyes and mouth from the face by using HOG and SVM leads to significantly improved results. The HOG is a good feature descriptor for specific regions and SVM is very suitable for the classification of these regions. This concludes that omitting some regions in the face has an impact on the accuracy performance. The expansion of the new form of datasets led to increasing the classification accuracy since the samples used to train data for the CNN have been increased. Where the accuracy of the suggested work report is about 94% for KDEF, 97% for JAFFE, and 97% for FER2013. In addition, the voting process is added to optimize the outcomes of the prediction which reach more than 98% for all utilized databases.

REFERENCES

- [1] González-Lozoya, S.M., de la Calleja, J., Pellegrin, L., Escalante, H.J., Medina, M.A., Benitez-Ruiz, A. (2020). Recognition of facial expressions based on CNN features. *Multimedia Tools and Applications*, 79: 13987-14007. <https://doi.org/10.1007/s11042-020-08681-4>
- [2] Darwin, C., Prodger, P. (1998). *The expression of the emotions in man and animals*. Oxford University Press, USA.
- [3] Tian, Y.I., Kanade, T., Cohn, J.F. (2001). Recognizing action units for facial expression analysis. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2): 97-115. <https://doi.org/10.1109/34.908962>
- [4] Williams, M.H., Broadley, S.A. (2008). SUNCT and SUNA: Clinical features and medical treatment. *Journal of Clinical Neuroscience*, 15(5): 526-534. <https://doi.org/10.1016/j.jocn.2006.09.006>
- [5] Xu, Y., Yang, W., Yu, X., Li, H., Cheng, T., Lu, X., Wang, Z.L. (2021). Real-time monitoring system of automobile driver status and intelligent fatigue warning based on triboelectric nanogenerator. *ACS Nano*, 15(4): 7271-7278. <https://doi.org/10.1021/acsnano.1c00536>
- [6] Laucka, A., Andriukaitis, D. (2015). Research of the defects in anesthetic masks. *Radioengineering: Proceedings of Czech and Slovak Technical Universities*, 24(4): 1033-1043. <https://doi.org/10.13164/re.2015.1033>
- [7] Zhao, Q., Jiang, J., Lei, Z., Yi, J. (2021). Detection method of eyes opening and closing ratio for driver's fatigue monitoring. *IET Intelligent Transport Systems*, 15(1): 31-42. <https://doi.org/10.1049/itr2.12002>
- [8] Søråa, R., Fostervold, M.E. (2021). Social domestication of service robots: The secret lives of Automated Guided Vehicles (AGVs) at a Norwegian hospital. *International Journal of Human-Computer Studies*, 152: 102627. <https://doi.org/10.1016/j.ijhcs.2021.102627>
- [9] Song, Y., Luximon, L.A., Luximon, Y. (2021). The effect of facial features on facial anthropomorphic trustworthiness in social robots. *Applied Ergonomics*, 94: 103420. <https://doi.org/10.1016/j.apergo.2021.103420>
- [10] Han, R., Koo, S.H. (2021). Big data analyses on key terms of wearable robots in social network services. *International Journal of Clothing Science and Technology*, 34(2): 285-298. <https://doi.org/10.1108/IJCST-11-2020-0180>
- [11] Glowacz, A. (2021). Fault diagnosis of electric impact drills using thermal imaging. *Measurement*, 171: 108815. <https://doi.org/10.1016/j.measurement.2020.108815>
- [12] Chen, X., Cao, M., Wei, H., Shang, Z., Zhang, L. (2021). Patient emotion recognition in human computer interaction system based on machine learning method and interactive design theory. *Journal of Medical Imaging and Health Informatics*, 11(2): 307-312. <https://doi.org/10.1166/jmih.2021.3293>
- [13] Paton, C., Kushniruk, A.W., Borycki, E.M., English, M., Warren, J. (2021). Improving the usability and safety of digital health systems: The role of predictive human-computer interaction modeling. *Journal of Medical Internet Research*, 23(5): e25281. <https://doi.org/10.2196/25281>
- [14] Ekman, P., Friesen, W.V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2): 124-129. <https://doi.org/10.1037/h0030377>
- [15] Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin*, 115(2): 268-287. <https://doi.org/10.1037/0033-2909.115.2.268>
- [16] Matsumoto, D. (1992). More evidence for the universality of a contempt expression. *Motivation and Emotion*, 16(4): 363-368. <https://doi.org/10.1007/BF00992972>

- [17] Dixit, A.N., Kasbe, T. (2020). A survey on facial expression recognition using machine learning techniques. In 2nd International Conference on Data, Engineering and Applications (IDEA). IEEE, 1-6. <https://doi.org/10.1109/IDEA49133.2020.9170706>
- [18] Mahrishi, M., Hiran, K.K., Meena, G. and Sharma, P. (2020). Machine learning and deep learning in real-time applications. IGI Global.
- [19] Górriz, J.M., Ramírez, J., Ortíz, A., Martínez-Murcia, F.J., Segovia, F., Suckling, J., Leming, M., Zhang, Y.D., Álvarez-Sánchez, J.R., Bologna, G. and Bonomini, P., Casado, F.E., Charte, D., Charte, F., Contreras, R., Cuesta-Infante, A., Duro, R.J., Fernández-Caballero, A., Fernández-Jover, E., Gómez-Vilda, P., Graña, M., Herrera, F., Iglesias, R., Lekova, A., de Lope, J., López-Rubio, E., Martínez-Tomás, R., Molina-Cabello, M.A., Montemayor, A.S., Novais, P., Palacios-Alonso, D., Pantrigo, J.J., Payne, B.R., de la Paz López, F., Pinninghoff, M.A., Rincón, M., Santos, J., Thurnhofer-Hemsi, K., Tsanas, A., Varela, R., Ferrández J.M. (2020). Artificial intelligence within the interplay between natural and artificial computation: Advances in data science, trends and applications. *Neurocomputing*, 410: 237-270. <https://doi.org/10.1016/j.neucom.2020.05.078>
- [20] Umer, S., Rout, R.K., Pero, C., Nappi, M. (2022). Facial expression recognition with trade-offs between data augmentation and deep learning features. *Journal of Ambient Intelligence and Humanized Computing*, 13(2): 721-735. <https://doi.org/10.1007/s12652-020-02845-8>
- [21] Podder, T., Bhattacharya, D., Majumdar, A. (2022). Dew computing-inspired mental health monitoring system framework powered by a lightweight CNN. In *Disruptive Technologies for Big Data and Cloud Applications: Proceedings of ICBDDCC 2021*. Singapore: Springer Nature Singapore, 905: 309-319. https://doi.org/10.1007/978-981-19-2177-3_31
- [22] Zhang, D., Tian, Q. (2021). A novel fuzzy optimized CNN-RNN method for facial expression recognition. *Elektronika Ir Elektrotechnika*, 27(5): 67-74. <https://doi.org/10.5755/j02.eic.29648>
- [23] Sadik, R., Anwar, S., Reza, M.L. (2021). Autismnet: recognition of autism spectrum disorder from facial expressions using mobilenet architecture. *International Journal of Advanced Trends in Computer Science and Engineering*, 10(1): 327-334. <https://doi.org/10.30534/ijatcse/2021/471012021>
- [24] Haque, M.I.U., Valles, D. (2018). A facial expression recognition approach using DCNN for autistic children to identify emotions. In 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 546-551. <https://doi.org/10.1109/IEMCON.2018.8614802>
- [25] Dalal, N., Triggs, B. (2005). Histograms of oriented gradients for human detection. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 1: 886-893. <https://doi.org/10.1109/CVPR.2005.177>
- [26] Zhong, Y., Sun, L., Ge, C., Fan, H. (2021). HOG-ESRs face emotion recognition algorithm based on HOG feature and ESRs method. *Symmetry*, 13(2): 1-18. <https://doi.org/10.3390/sym13020228>
- [27] Dadi, H.S., Pillutla, G.M. (2016). Improved face recognition rate using HOG features and SVM classifier. *IOSR Journal of Electronics and Communication Engineering*, 11(04): 34-44. <https://doi.org/10.9790/2834-1104013444>
- [28] Donia, M.M., Youssif, A.A., Hashad, A. (2014). Spontaneous facial expression recognition based on histogram of oriented gradients descriptor. *Computer and Information Science*, 7(3): 31-37. <https://doi.org/10.5539/cis.v7n3p31>
- [29] Sakthivel, K., Nallusamy, R., Kavitha, C. (2015). Color image segmentation using SVM pixel classification image. *International Journal of Computer and Information Engineering*, 8(10): 1924-1930.
- [30] Baghaee, H.R., Mlakić, D., Nikolovski, S., Dragicević, T. (2020). Support vector machine-based islanding and grid fault detection in active distribution networks. In *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 8(3): 2385-2403. <https://doi.org/10.1109/JESTPE.2019.2916621>
- [31] Soffer, S., Ben-Cohen, A., Shimon, O., Amitai, M.M., Greenspan, H., Klang, E. (2019). Convolutional neural networks for radiologic images: A radiologist's guide. *Radiology*, 290(3): 590-606. <https://doi.org/10.1148/radiol.2018180547>
- [32] Antonopoulou, A., Balasis, G., Papadimitriou, C., Boutsis, A.Z., Rontogiannis, A., Koutroumbas, K., Daglis, I.A., Giannakis, O. (2022). Convolutional neural networks for automated ULF wave classification in swarm time series. *Atmosphere*, 13(9): 1488. <https://doi.org/10.3390/atmos13091488>
- [33] Krizhevsky, A., Sutskever, I., Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25(2): 1097-1105.